

Control of nonlinear systems: a model inversion approach

C. Novara, M. Milanese

Abstract—A novel control design approach for general nonlinear systems is presented in this paper. The approach is based on the identification of a polynomial model of the system to control and on the on-line inversion of this model. An efficient technique is developed to perform the inversion, which allows an effective control implementation on real-time processors.

I. INTRODUCTION

Consider a nonlinear discrete-time SISO system in regression form:

$$\begin{aligned} y_{t+1} &= g(\mathbf{y}_t, \mathbf{u}_t, \boldsymbol{\xi}_t) \\ \mathbf{y}_t &= (y_t, \dots, y_{t-n+1}) \\ \mathbf{u}_t &= (u_t, \dots, u_{t-n+1}) \\ \boldsymbol{\xi}_t &= (\xi_t, \dots, \xi_{t-n+1}) \end{aligned} \quad (1)$$

where $u_t \in U \subset \mathbb{R}$ is the input, $y_t \in \mathbb{R}$ is the output, $\boldsymbol{\xi}_t \in \Xi \subset \mathbb{R}^{n_\xi}$ is a disturbance including both process and measurement noises, and n is the system order. U and Ξ are compact sets. In particular, $U \doteq [\underline{u}, \bar{u}]$ accounts for input saturation.

Suppose that the system (1) is unknown, but a set of measurements is available:

$$\mathcal{D} \doteq \{\tilde{u}_t, \tilde{y}_t\}_{t=1-L}^0 \quad (2)$$

where \tilde{u}_t and \tilde{y}_t are bounded for all $t = 1 - L, \dots, 0$. The tilde is used to indicate the input and output samples of the data set (2).

Let $\mathcal{Y}^0 \subseteq \mathbb{R}^n$ be a set of initial conditions of interest for the system (1) and, for a given initial condition $\mathbf{y}_0 \in \mathcal{Y}^0$, let $\mathcal{Y}(\mathbf{y}_0) \subseteq \ell_\infty$ be a set of output sequences of interest.

The aim is to control the system (1) in such a way that, starting from any initial condition $\mathbf{y}_0 \in \mathcal{Y}^0$, the system output sequence $\mathbf{y} = (y_1, y_2, \dots)$ tracks any reference sequence $\mathbf{r} = (r_1, r_2, \dots) \in \mathcal{Y}(\mathbf{y}_0)$. The set of all solutions of interest is defined as $\mathcal{Y} \doteq \{\mathcal{Y}(\mathbf{y}_0) : \mathbf{y}_0 \in \mathcal{Y}^0\}$. The set of all possible disturbance sequences is defined as $\Xi \doteq \{\boldsymbol{\xi} = (\xi_1, \xi_2, \dots) : \xi_t \in \Xi, \forall t\}$.

After a brief section regarding the notation used in the paper, an approach called NIC (Nonlinear Inversion Control) is proposed for the design of a controller K^{nl} , allowing the accomplishment of the above task.

II. NOTATION

A column vector $\mathbf{x} \in \mathbb{R}^{n_x \times 1}$ is denoted as $\mathbf{x} = (x_1, \dots, x_{n_x})$. A row vector $\mathbf{x} \in \mathbb{R}^{1 \times n_x}$ is denoted as

$\mathbf{x} = [x_1, \dots, x_{n_x}] = (x_1, \dots, x_{n_x})^\top$, where \top indicates the transpose.

A discrete-time signal (i.e. a sequence of vectors) is denoted with the bold style: $\mathbf{x} = (x_1, x_2, \dots)$, where $x_t \in \mathbb{R}^{n_x \times 1}$ and $t = 1, 2, \dots$ indicates the discrete time; $x_{i,t}$ is the i th component of the signal \mathbf{x} at time t .

A regressor, i.e. a vector that, at time t , contains n present and past values of a variable, is indicated with the bold style and the time index: $\mathbf{x}_t = (x_t, \dots, x_{t-n+1})$.

The ℓ_p norms of a vector $\mathbf{x} = (x_1, \dots, x_{n_x})$ are defined as

$$\|\mathbf{x}\|_p \doteq \begin{cases} (\sum_{i=1}^{n_x} |x_i|^p)^{\frac{1}{p}}, & p < \infty, \\ \max_i |x_i|, & p = \infty. \end{cases}$$

The ℓ_∞ norm is also used to denote the absolute value of a scalar: $\|\mathbf{x}\|_\infty \equiv |x|$ for $x \in \mathbb{R}$.

The ℓ_p norms of a signal $\mathbf{x} = (x_1, x_2, \dots)$ are defined as

$$\|\mathbf{x}\|_p \doteq \begin{cases} (\sum_{t=1}^{\infty} \sum_{i=1}^{n_x} |x_{i,t}|^p)^{\frac{1}{p}}, & p < \infty, \\ \max_{i,t} |x_{i,t}|, & p = \infty, \end{cases}$$

where $x_{i,t}$ is the i th component of the signal \mathbf{x} at time t . These norms give rise to the well-known ℓ_p Banach spaces.

III. NIC CONTROL DESIGN

The proposed approach relies on identifying from the data (2) a model of the form

$$\begin{aligned} \hat{y}_{t+1} &= f(\mathbf{y}_t, \mathbf{u}_t) \equiv f(\mathbf{q}_t, u_t) \\ \mathbf{q}_t &= (\mathbf{y}_t, \mathbf{u}_t) = (y_t, \dots, y_{t-n+1}, u_t, \dots, u_{t-n+1}) \end{aligned} \quad (3)$$

where u_t and y_t are the system input and output, and \hat{y}_t is the model output. For simplicity, the model is supposed of the same order as the system but this choice is not necessary: all the results presented in the paper hold also when the model and system orders are different. Indications on the choice of the model order are given in Section IV.

A parametric structure is taken for the function f :

$$f(\mathbf{q}_t, u_t) = \sum_{i=1}^N \alpha_i \phi_i(\mathbf{q}_t, u_t) \quad (4)$$

where ϕ_i are basis functions and α_i are parameters to be identified. The basis function choice is in general a crucial step, [1], [2], [3]. In the present NIC approach, polynomial functions are used. The motivations are mainly two: (1) polynomials have been proved to be effective approximators in a huge number of problems; (2) as we will see later, they allow a “fast” controller evaluation. The identification of the parameter vector $\boldsymbol{\alpha} \doteq (\alpha_1, \dots, \alpha_N)$ can be performed by means of convex optimization, as shown in Section IV.

Once a model of the form (3) has been identified, the controller K^{nl} is obtained by its inversion:

Suppose that, at a time $t > 0$, the reference value for the time $t + 1$ is r_{t+1} and the current regressor is \mathbf{q}_t . Inversion consists in finding a command input u_t^{nl} such that the model output at time $t + 1$ is “close” to r_{t+1} :

$$\hat{y}_{t+1} = f(\mathbf{q}_t, u_t^{nl}) \cong r_{t+1}. \quad (5)$$

The latter equality may be not exact for two reasons: (1) no $u_t^{nl} \in U$ may exist for which \hat{y}_{t+1} is exactly equal to r_{t+1} ; (2) values of u_t^{nl} with a limited ℓ_2 norm may be of interest, in order to have a not too high command activity. This kind of inversion is called (approximate) right-inversion and can be performed also when f is not bijective with respect to u_t (e.g., for some r_{t+1} and \mathbf{q}_t , more than one value of u_t may exist such that (5) holds).

The command input u_t^{nl} yielding (5) can be computed according to the following optimality criterion:

$$u_t^{nl} = \arg \min_{u \in U} J(u) \quad (6)$$

subject to $u \in U$.

The objective function is given by

$$J(u) = \frac{1}{\rho_y} (r_{t+1} - f(\mathbf{q}_t, u))^2 + \frac{\mu}{\rho_u} u^2 \quad (7)$$

where $\rho_y \doteq \|(\tilde{y}_{1-L}, \dots, \tilde{y}_0)\|_2^2$ and $\rho_u \doteq \|(\tilde{u}_{1-L}, \dots, \tilde{u}_0)\|_2^2$ are normalization constants computed from the data set (2), and $\mu \geq 0$ is a design parameter, allowing us to determine the trade-off between tracking precision and command activity. Indications on the choice of μ are given in Section IV.

Note that the objective function (7) is in general non-convex. Moreover, the optimization problem (6) has to be solved on-line, and this may require a long time compared to the sampling time used in the application of interest. In order to overcome these two relevant problems, a technique is now proposed, allowing a very efficient computation of the optimal command input u_t^{nl} .

Since a polynomial basis function expansion has been considered for f , it follows that the objective function $J(u)$ is a polynomial in u . The minima of $J(u)$ can thus be found considering the roots of its derivative: Define the set

$$U^s \doteq \left(\text{Roots} \left(\frac{dJ(u)}{du} \right) \cap U \right) \cup \{\underline{u}, \bar{u}\}$$

where $\text{Roots}(\cdot)$ denotes the set of all real roots of \cdot , and \underline{u} and \bar{u} are the boundaries of U . The optimal command input is given by

$$u_t^{nl} = K^{nl}(r_{t+1}, \mathbf{q}_t) \doteq \arg \min_{u \in U^s} J(u) \quad (8)$$

where it has been considered that U^s depends on the reference r_{t+1} and regressor \mathbf{q}_t .

The nonlinear controller K^{nl} is fully defined by the control law (8).

Remark 1: The derivative $dJ(u)/du$ can be computed analytically. Moreover, U^s is composed by a “small” number of elements:

$$\text{card}(U^s) < \deg(J(u)) + 2$$

where card is the set cardinality and \deg indicates the polynomial degree. The evaluation of u_t^{nl} through (8) is thus extremely fast, since it just requires to find the real roots of a polynomial whose analytical expression is known and to compute the objective function for a “small” number of values. This fact allows a very efficient controller implementation on real-time devices. ■

Remark 2: The system (1) is not required to be stable and in general no preliminary stabilizing controllers are needed. The only guideline is to generate the data using input signals for which the system output does not diverge. This can be easily done for many nonlinear systems like the single-corner model considered below. Indeed, many systems are characterized by trajectories that are unstable but bounded (a typical feature of chaotic systems). In the presence of unbounded trajectories, for which a suitable input signal can hardly be found, a preliminary stabilizing controller may be required. The preliminary controller can also be a human operator, who is able to drive the system within a bounded domain, see [4], [5]. ■

IV. MODEL IDENTIFICATION ALGORITHM

In this section, the algorithm for identifying the model (3), required by the nonlinear controller K^{nl} , is proposed. Systematic criteria for the choice of the involved identification/design parameters are also provided.

Choose a set of polynomial basis functions ϕ_i . For example, these functions can be generated as products of univariate polynomials, where the independent variables are scaled to range in the interval $[-1, 1]$. In most cases, no large polynomial degrees are required: we observed in several simulated and real-world applications that a degree $\lesssim 8$ is sufficient to guarantee satisfactory model accuracy and control performance.

Define

$$\tilde{\mathbf{y}} \doteq (\tilde{y}_{t_1+1}, \dots, \tilde{y}_{t_2+1})$$

$$\Phi \doteq \begin{bmatrix} \phi_1(\tilde{\mathbf{y}}_{t_1}, \tilde{\mathbf{u}}_{t_1}) & \cdots & \phi_N(\tilde{\mathbf{y}}_{t_1}, \tilde{\mathbf{u}}_{t_1}) \\ \vdots & \ddots & \vdots \\ \phi_1(\tilde{\mathbf{y}}_{t_2}, \tilde{\mathbf{u}}_{t_2}) & \cdots & \phi_N(\tilde{\mathbf{y}}_{t_2}, \tilde{\mathbf{u}}_{t_2}) \end{bmatrix}$$

where $t_1 \doteq 1 - L + n$, $t_2 \doteq -1$, and \tilde{u}_t and \tilde{y}_t are the input-output measurements of the data set (2). Consider the set $SC \subset \mathbb{R}^N$, defined as

$$SC(\gamma_y, \eta, \rho) \doteq \{\beta : |\tilde{y}_{l+1} - \tilde{y}_{k+1} + (\Phi_k - \Phi_l)\beta| \leq \gamma_y \rho \|\tilde{\mathbf{y}}_l - \tilde{\mathbf{y}}_k\|_\infty + 2\eta\rho, k \in \mathcal{T}, l \in \Upsilon_k\}$$

where $\mathcal{T} \doteq \{t_1, \dots, t_2\}$ and Υ_k is the set of indexes given by

$$\Upsilon_k \doteq \{i : \|\tilde{\mathbf{u}}_k - \tilde{\mathbf{u}}_i\|_\infty \leq \zeta\}$$

and ζ is the minimum value for which every set Υ_k contains at least two elements. SC is defined by a set of linear inequalities in β and is thus convex in β .

The parameter vector $\alpha \doteq (\alpha_1, \dots, \alpha_N)$ of the model defined by (3) and (4) can be identified by means of the following algorithm, completely based on convex optimization. Note

that the algorithm is “self-tuning”, in the sense that all the required parameters are chosen by the algorithm itself, without requiring extensive trial and error procedures.

Algorithm 1

Initialization: choose a “low” model order (e.g. $n = 1$).

1) Construct the vector $\tilde{\mathbf{y}}$ and matrix Φ as indicated above.

2) Compute

$$\eta = \min_{\beta \in \mathbb{R}^N} \|\tilde{\mathbf{y}} - \Phi\beta\|_\infty.$$

3) Solve the optimization problem

$$\begin{aligned} \alpha &= \arg \min_{\beta \in \mathbb{R}^N} \|\beta\|_1 \\ \text{subject to} \\ (a) \quad &\beta \in SC(\gamma_y, \eta, \rho) \\ (b) \quad &\|\tilde{\mathbf{y}} - \Phi\beta\|_\infty \leq \eta\rho \end{aligned} \quad (9)$$

where Φ_k denotes the k th row of the matrix Φ , γ_y is the minimum value for which the constraints (a) and (b) are feasible, and ρ is a real number slightly larger than 1.

4) Repeat steps 1-3 for increasing model order. Stop when no significant reductions of γ_y are observed.

5) Repeat steps 3-4 for increasing ρ . Stop when $\gamma_y < 1$. ■

The algorithm allows the achievement of three important features:

1) **Closed-loop stability.** As proven in [4], under reasonable conditions, constraint (a) ensures that the function $\Delta \doteq g - f$ has a Lipschitz constant wrt \mathbf{y}_t non larger than γ_y , as $L \rightarrow \infty$. On the other hand, a theoretical analysis shows that having this constant smaller than 1 is a key condition for closed-loop stability. Another required condition is that $\Gamma_y < 1 - \gamma_y$, where Γ_y is the input-output gain of the system formed by the cascade connection of the controller and the model (this latter working in prediction). Since Γ_y can be imposed arbitrarily (see the discussion below), it can be concluded that *Algorithm 1 is able to ensure closed-loop stability when the number of data becomes large.*

2) **“Small” tracking error.** Constraint (b) is aimed at providing a model with a “small” prediction error (this error, evaluated on the design data set, is given by $\|\tilde{\mathbf{y}} - \Phi\alpha\|_\infty$). According to the mentioned theoretical analysis, reducing this error allows us to obtain a “small” tracking error. Note that there is a trade-off between stability and tracking performance: In step 5, ρ is increased until the stability condition is met. However, increasing ρ causes an increase of the prediction error and, consequently, of the tracking error.

3) **Model sparsity.** In step 3, the ℓ_1 norm of the coefficient vector β is minimized, leading to a sparse coefficient vector α , i.e. a vector with a “small” number of non-zero elements, [6], [7], [8], [9]. Sparsity is important to

ensure a low complexity and a high regularity of the model, limiting at the same time well known issues such as over-fitting and the curse of dimensionality. Sparsity allows also an efficient implementation of the model/controller on real-time processors, which may have limited memory and computation capacities. ■

Once a model has been identified, the nonlinear controller K^{nl} is obtained by its inversion, as explained in Section III. Only one design parameter needs to be chosen for this inversion: the weight μ in (7). If no particular requirements on the activity of the command input u_t have to be satisfied, the simplest choice is $\mu = 0$. Otherwise, if the input activity has to be reduced, a value $0 < \mu \leq \bar{\mu}$ can be chosen, where $\bar{\mu}$ is the maximum value for which the stability condition $\Gamma_y < 1 - \gamma_y$ holds. Γ_y is the input-output gain of the system formed by the cascade connection of the controller and the model. This condition can be checked (approximately) by deriving an estimate $\hat{\Gamma}_y$ of Γ_y from the data (2). Let

$$\mathcal{D}^\Gamma \doteq \{\tilde{w}_t, \hat{y}_{t+1}\}_{t=1-L+m}^{-1} \quad (10)$$

where

$$\begin{aligned} \hat{y}_t &= f(\tilde{\mathbf{y}}_{t-1}, \tilde{\mathbf{u}}_{t-1}^{nl}) \\ \tilde{\mathbf{u}}_{t-1}^{nl} &= K^{nl}(\tilde{\mathbf{y}}_t, \tilde{\mathbf{q}}_{t-1}) \\ \tilde{\mathbf{q}}_{t-1} &= (\tilde{y}_{t-1}, \dots, \tilde{y}_{t-n}, \tilde{u}_{t-2}^{nl}, \dots, \tilde{u}_{t-n}^{nl}) \\ \tilde{w}_t &= (\tilde{y}_t, \dots, \tilde{y}_{t-m+1}), \end{aligned} \quad (11)$$

\tilde{u}_t and \tilde{y}_t are the input-output measurements of the data set (2), and $m \gg n$. The estimate $\hat{\Gamma}_y$ can be obtained applying the validation method of [10] to the data set (10) (the method is summarized in the Appendix). Observing that $\tilde{u}_{t-1}^{nl} \equiv \tilde{u}_{t-1}^{nl}(\mu)$ and thus $\hat{\Gamma}_y \equiv \hat{\Gamma}_y(\mu)$, μ must be chosen in such a way that $\hat{\Gamma}_y(\mu) < 1 - \gamma_y$.

Remark 3: The stability conditions $\gamma_y < 1$ and $\Gamma_y < 1 - \gamma_y$ can give indications on the choice of the control system sampling time T_s : As discussed in [11], a too small T_s leads to models where $\hat{y}_{t+1} \cong y_t$. These kinds of models have a strong dependence on past outputs and a weak dependence on the input, resulting in large values of γ_y and Γ_y . It is thus expected that γ_y and Γ_y can be reduced by increasing T_s . Clearly, to capture the relevant dynamics of the system and allow a prompt control action, T_s must be not “too large”. ■

V. APPENDIX: NONLINEAR SET MEMBERSHIP VALIDATION PROCEDURE

In this appendix, the validation method of [10] is summarized, suitably adapted for the present setting. This method is useful within the NIC approach for estimating the constant Γ_y appearing in Sections IV and I.

Suppose that Γ_y is the Lipschitz constant of an unknown function \mathbf{f} . Let a set of data $(\tilde{w}_t, \hat{y}_{t+1})$, $t \in \mathcal{T}$ be available, described by

$$\hat{y}_{t+1} = \mathbf{f}(\tilde{w}_t) + d_t, \quad t \in \mathcal{T}$$

where \mathcal{T} is a suitable set of indexes and d_t is a noise. This noise may also include errors due to the fact that \mathbf{f} is not

Lipschitz continuous (e.g. in the case where f is the sum of a Lipschitz continuous function plus a discontinuous but bounded function). Assume that $d_t \in B_\varepsilon$, where B_ε is the ℓ_∞ ball with radius ε , and that $f \in \mathcal{F}(\Gamma_y)$, where $\mathcal{F}(\Gamma_y)$ is the set of Lipschitz continuous functions on the domain of \tilde{w}_t with constant Γ_y . Under this assumption, we have that $f \in FFS$, where FFS is the Feasible Function Set.

Definition 1: Feasible Function Set:

$$FFS \doteq \{f \in \mathcal{F}(\Gamma_y) : \hat{y}_{t+1} - f(\tilde{w}_t) \in B_\varepsilon, t \in \mathcal{T}\}. \quad \blacksquare$$

According to this definition, FFS is the set of all functions consistent with prior assumptions and data. As typical in any identification/estimation theory, the problem of checking the validity of prior assumptions arises. The only thing that can be actually done is to check if prior assumptions are invalidated by the data, evaluating if no function exists consistent with data and assumptions, i.e. if FFS is empty.

Definition 2: Prior assumptions are validated if $FFS \neq \emptyset$. \blacksquare

The following result provides necessary and sufficient conditions for prior assumption validation. Define the function $\bar{f}(\Gamma_y, w) \doteq \min_{t \in \mathcal{T}} (\hat{y}_{t+1} + \varepsilon + \Gamma_y \|w - \tilde{w}_t\|)$.

Theorem 1: A sufficient condition for prior assumptions to be validated is:

$$\bar{f}(\Gamma_y, \tilde{w}_t) > \hat{y}_{t+1} - \varepsilon, t \in \mathcal{T}.$$

Proof. See Theorem 1 in [10]. \blacksquare

The validation Theorem 1 can be used for assessing the value of the Lipschitz constant Γ_y so that the sufficient condition holds. Suppose that ε has been chosen by means of any criterion (e.g. based on some prior knowledge on the noises, or by means of standard filtering/smoothing techniques, or also using the dispersion function defined in [12]). The constant

$$\Gamma_y^{\min} \doteq \inf_{\bar{f}(\Gamma, \tilde{w}_t) > \hat{y}_{t+1} - \varepsilon, t \in \mathcal{T}} \Gamma \quad (12)$$

represents the minimum Lipschitz constant for which the prior assumptions are validated. A reasonable estimate of Γ_y is thus a value slightly larger than Γ_y^{\min} . Note that the evaluation of Γ_y^{\min} is quite simple, as shown by the examples in [10] and [13].

REFERENCES

- [1] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Glorennec, H. Hjalmarsson, and A. Juditsky, "Nonlinear black-box modeling in system identification: a unified overview," *Automatica*, vol. 31, pp. 1691–1723, 1995.
- [2] K. Hsu, C. Novara, T. Vincent, M. Milanese, and K. Poolla, "Parametric and nonparametric curve fitting," *Automatica*, vol. 42/11, pp. 1869–1873, 2006.
- [3] C. Novara, T. Vincent, K. Hsu, M. Milanese, and K. Poolla, "Parametric identification of structured nonlinear systems," *Automatica*, vol. 47, no. 4, pp. 711 – 721, 2011.
- [4] C. Novara, L. Fagiano, and M. Milanese, "Direct feedback control design for nonlinear systems," *Automatica*, vol. 49, no. 4, pp. 849–860, 2013.
- [5] L. Fagiano and C. Novara, "Learning controllers from data: control of a power kite used for wind energy conversion," http://lorenzofagiano.altervista.org/movies/LC_kite_LD.mp4, 2012.
- [6] J. Fuchs, "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3601 –3608, oct. 2005.
- [7] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse over-complete representations in the presence of noise," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6 – 18, jan. 2006.
- [8] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489 – 509, feb. 2006.
- [9] J. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030 –1051, mar. 2006.
- [10] M. Milanese and C. Novara, "Set membership identification of nonlinear systems," *Automatica*, vol. 40/6, pp. 957–975, 2004.
- [11] G. Goodwin, J. Yuz, J. Aguero, and M. Cea, "Sampling and sampled-data models," in *American Control Conference, plenary lecture*, Baltimore, MD, USA, 2010.
- [12] K. Hsu, T. Vincent, C. Novara, M. Milanese, and K. Poolla, "Identification of nonlinear maps in interconnected systems," in *44th IEEE Conference on Decision and Control and European Control Conference*, Seville, Spain, 2005.
- [13] M. Milanese and C. Novara, "Set membership prediction of nonlinear time series," *IEEE Transactions on Automatic Control*, vol. 50, no. 11, pp. 1655–1669, 2005.